[TRANSLATION]

PCT/JP02/13053

AMENDMENT UNDER ARTICLE 34 PCT

filed August 7, 2003

context to recognize a continuous input speech, comprising an acoustic analysis section analyzing the input speech to obtain feature parameter time series; a word lexicon in which each of words included in vocabulary is stored in a form of a sub-word network or in a sub-word tree structure; a language model storage unit in which language models representing information regarding connection between words is stored; a context dependent acoustic model storage unit in which the context dependent acoustic models are stored in a form of sub-word state trees in each of which state sequences of a plurality of sub-word models of the context dependent acoustic models are organized in a tree structure; a matching unit developing hypotheses of sub-words by referencing the sub-word state tree representing the context dependent acoustic models, the word lexicon and the language models, and performing matching between the feature parameter time series and the developed hypotheses so as to output, as a word lattice, word information including a word, an accumulated score and a beginning start frame with respect to a hypothesis representing a word end portion; and a search unit for searching the word lattice to generate recognition results.

[0013] According to the above constitution, sub-word hypotheses are developed by referring to the sub-word state trees formed by placing the context dependent acoustic

models dependent on the sub-word context in a tree structure, the word lexicon and the language model. Therefore, what is necessary is only to develop one hypothesis regardless of a head or leading sub-word of the next word, which allows drastic decrease of a total number of states in all the hypotheses. More specifically, it becomes possible to significantly reduce the hypothesis developing amount and easily develop hypotheses regardless of in-word or word-boundary state. Further, the matching unit allows significant reduction of the amount of operation when the feature parameter series from the acoustic analysis section are matched with the developed hypotheses.

[0014] In one embodiment, the context dependent acoustic models stored in the context dependent acoustic model storage unit (3) are context dependent acoustic models in which a center sub-word depends on sub-words preceding and succeeding the center sub-word respectively, and the state sequences of sub-word models having identical preceding sub-words and identical center sub-words are organized in a tree structure.

[0015] According to this embodiment, the hypotheses are developed by using the sub-word state trees formed by placing the state sequences of the sub-word models having the same preceding sub-word and the same center sub-word in a tree structure. Therefore, when developing the next

hypothesis, attention should be paid only to a center sub-word in the preceding or end hypothesis and a sub-word state tree having a corresponding preceding sub-word should be developed. More precisely, even with the presence of a multiplicity of succeeding sub-words, the number of hypotheses to be developed can be smaller, so that the hypotheses can be developed easily.

[0016] In one embodiment, the context dependent acoustic models are state sharing models in which a plurality of sub-word models share states.

[0017] According to this embodiment, state sharing by a plurality of sub-word models makes it possible to combine the shared states together when placed in a tree structure, thereby allowing decrease of the number of nodes. Therefore, the processing amount during matching operation by the matching unit can be reduced significantly.

[0018] In one embodiment, when developing the hypotheses by referencing the sub-word state tree, the matching unit puts a flag on states connectable to each other in the sub-word state trees that represent the hypotheses, by using information on connectable sub-words obtained from the word lexicon and the language model.

[0019] According to this embodiment, of the states in the sub-word state tree constituting the developed hypothesis, states connectable to each other are flagged. This limits

the states that require Viterbi calculation during matching operation, thereby allowing further decrease of the matching amount.

[0020] In one embodiment, during a matching operation, the matching unit calculates scores of the developed hypotheses based on the feature parameter time series, and prunes the hypotheses in conformity to criteria including a threshold value of the scores or a quantity of hypotheses.

[0021] According to this embodiment, the hypothesis pruning is performed during the matching operation, so that hypotheses with low likelihood to be a word or words are deleted, which allows significant reduction of the following matching operation amount.

[0022] The present invention also provides a continuous speech recognition method which uses, as a recognition unit, a sub-word determined depending on an adjacent sub-word and which uses context dependent acoustic models dependent on sub-word context to recognize a continuous input speech, comprising analyzing the input speech to obtain feature parameter time series by an acoustic analysis section; developing hypotheses of sub-words by referencing a sub-word state tree formed by placing state sequences of the context dependent acoustic models in a tree structure, a word lexicon describing each of words included in vocabulary in a form of a sub-word network or in a sub-word tree structure,

and a language model representing information regarding connection between words, and performing matching between the feature parameter time series and the developed hypotheses so as to generate, as a word lattice, word information including a word, an accumulated score and a beginning start frame with respect to a hypothesis regarding a word end portion, by a matching unit; and searching the word lattice to generate recognition results by a search unit.

[0023] According to the above constitution, as with the case of the continuous speech recognition apparatus of the invention, hypotheses are developed by referring to the sub-word state tree formed by placing the context dependent acoustic models in a tree structure. Therefore, what is necessary is only to develop one hypothesis regardless of the head sub-word of the succeeding word, which makes it possible to easily develop hypotheses regardless of in-word or word-boundary state. Further, the amount of matching operation to be done for matching between the feature parameter series and the developed hypotheses is significantly reduced.

[0024] A continuous speech recognition program according to the present invention makes a computer function as the acoustic analysis section, the word lexicon, the language model storage unit, the context dependent acoustic model

storage unit, the matching unit, and the search unit in the continuous speech recognition device of the present invention.

[0025] According to the above constitution, as with the case of the continuous speech recognition apparatus of the invention, only one hypothesis may be developed regardless of the leading sub-word of the succeeding word, which makes it possible to easily develop hypotheses regardless of in-word or word-boundary state. Further, the amount of matching operation to be done for matching between the feature parameter series and the developed hypotheses is significantly reduced.

[0026] A program recording medium according to the present invention has the continuous speech recognition program of the present invention stored therein.

[0027] According to the above constitution, as with the case of the continuous speech recognition apparatus of the invention, only one hypothesis may be developed regardless of the leading sub-word of the succeeding word, which makes it possible to easily develop hypotheses regardless of in-word or word-boundary state. Further, the amount of matching operation to be done for matching between the feature parameter series and the developed hypotheses is significantly reduced.

WHAT IS CLAIMED IS:

1. A continuous speech recognition apparatus which uses, as a recognition unit, a sub-word determined depending on an adjacent sub-word and which uses context dependent acoustic models dependent on sub-word context to recognize a continuous input speech, comprising:

an acoustic analysis section (1) analyzing the input speech to obtain feature parameter time series;

a word lexicon (4) in which each of words included in vocabulary is stored in a form of a sub-word network or in a sub-word tree structure;

a language model storage unit (5) in which language models representing information regarding connection between words is stored;

a context dependent acoustic model storage unit (3) in which the context dependent acoustic models are stored in a form of sub-word state trees in each of which state sequences of a plurality of sub-word models of the context dependent acoustic models are organized in a tree structure;

a matching unit (2) developing hypotheses of sub-words by referencing the sub-word state tree representing the context dependent acoustic models, the word lexicon (4) and the language models, and performing matching between the feature parameter time series and the developed hypotheses

so as to output, as a word lattice, word information including a word, an accumulated score and a beginning start frame with respect to a hypothesis representing a word end portion; and

a search unit (8) for searching the word lattice to generate recognition results.

2. The continuous speech recognition apparatus as defined in Claim 1, wherein

the context dependent acoustic models stored in the context dependent acoustic model storage unit (3) are context dependent acoustic models in which a center sub-word depends on sub-words preceding and succeeding the center sub-word respectively, and the state sequences of sub-word models having identical preceding sub-words and identical center sub-words are organized in a tree structure.

3. The continuous speech recognition apparatus as defined in Claim 2, wherein

the context dependent acoustic models are state sharing models in which a plurality of sub-word models share states.

4. The continuous speech recognition apparatus as defined in Claim 1, wherein

when developing the hypotheses by referencing the sub-word state tree, the matching unit (2) puts a flag on states connectable to each other in the sub-word state trees that represent the hypotheses, by using information on connectable sub-words obtained from the word lexicon (4) and the language model.

5. The continuous speech recognition apparatus as defined in Claim 1, wherein

during a matching operation, the matching unit (2) calculates scores of the developed hypotheses based on the feature parameter time series, and prunes the hypotheses in conformity to criteria including a threshold value of the scores or a quantity of hypotheses.

6. A continuous speech recognition method which uses, as a recognition unit, a sub-word determined depending on an adjacent sub-word and which uses context dependent acoustic models dependent on sub-word context to recognize a continuous input speech, comprising:

analyzing the input speech to obtain feature parameter time series by an acoustic analysis section;

developing hypotheses of sub-words by referencing a sub-word state tree formed by placing state sequences of the context dependent acoustic models in a tree structure, a

word lexicon describing each of words included in vocabulary in a form of a sub-word network or in a sub-word tree structure, and a language model representing information regarding connection between words, and performing matching between the feature parameter time series and the developed hypotheses so as to generate, as a word lattice, word information including a word, an accumulated score and a beginning start frame with respect to a hypothesis regarding a word end portion, by a matching unit; and

searching the word lattice to generate recognition results by a search unit.

7. A continuous speech recognition program that makes a computer function as the acoustic analysis section (1), the word lexicon (4), the language model storage unit (5), the context dependent acoustic model storage unit (3), the matching unit (2) and the search unit (8) as recited in Claim 1.

8. A program recording medium readable by computer, having the continuous speech recognition program as defined in Claim 7 stored therein.